Introduction to Transcription

Introduction

Transcription is the copying of DNA into a form that can then be translated into a protein. In prokaryotes, this process happens in the cytoplasm. In eukaryotes transcription happens in the nucleus, while translation is in the cytoplasm. The **primary transcript** is the exact **RNA copy** of the **DNA code**, with a U traded for every T.

There are three strands we need to pay attention to. The **coding strand** is the DNA that RNA polymerase never touches. The coding strand has the genetic code, in order, 5' to 3', that'll match the sequence of amino acids. It informs the ribosome of the amino acid sequence, but can't do that from within the nucleus. The **transcribing strand** is the mRNA transcript—it's the RNA copy of the DNA code. It must carry the code of the coding strand exactly. Therefore, the **transcribing strand** has the same orientation and same sequence as the coding strand.

The **template strand** is the strand that the RNA polymerase uses as a template. The template strand and the coding strand are each one side of double-stranded DNA, antiparallel and complementary to each other. The transcribing strand uses the template strand, so will be built complementary and antiparallel to the template strand. If both the coding strand and the transcribing strand are complementary and antiparallel to the template strand, then the coding and the transcribing strand are the same as each other!

Orientation and Making This Easy

DNA replication was hard because there were two replication forks with helicases moving in both directions. Then there was a DNA polymerase that could move only one direction, but had to fill in the spaces behind it. This was because both strands acted as templates simultaneously. DNA replication is meant to copy the entire genetic material, doubling it for cell division.

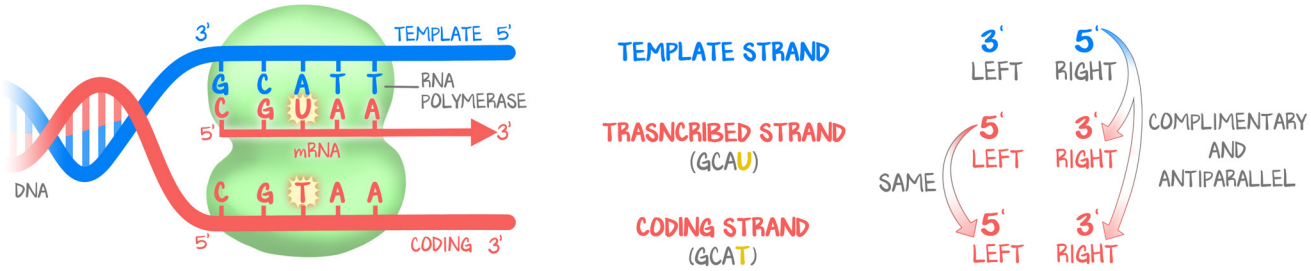

Figure 7.1: Coding Strand, Template Strand, and mRNA Transcribed Strand
Because the coding strand is complementary and antiparallel to the template strand, and the growing transcribed strand is complementary and antiparallel to the template strand, the coding strand and transcribing strand are the same. The only difference is that the mRNA is RNA, and so uses U instead of a T. Also, orientation 5' left 3' right ensures that the sequence that matters (coding and mRNA) is read left-to-right.

RNA transcription doesn't have any of that worry because in RNA transcription, only **one strand is copied**. But it's hard for different reason—because the direction matters. Since only one strand is copied, and only part of the strand at that, for the message to make sense it must be read with the correct **orientation**. Getting the orientation of the strands right is the most important piece of understanding the transcription process.

RNA polymerase adds a nucleotide to a growing strand at the 3' end. That is, it starts at the 5' end of the transcribing strand and moves towards its 3' end. That means it's starting at the 3' end of the template strand and moving towards the 5' end of the template strand. It's reading the template strand 3' to 5', but building the transcribing strand 5' to 3'.

This will get very difficult if you don't orient the DNA. Put the **mRNA 5' left, 3' right**. That means start the mRNA on the left side of the page, the first nucleotide on the left of the page, the end with a free 5-position carbon attached to a phosphate to the left. Farthest right on the page is the nucleotide most recently added to the 3' end of the transcribing strand.

Keep the DNA 5' left 3' right. Since we want to read naturally in English left to right, and the mRNA is 5' left 3' right, we should have our **coding strand** 5' left 3' right.

This means that the template strand must be 3' left, 5' right, the reverse of both the mRNA transcribing strand and the coding strand. We'll intentionally leave the **template strand out** of most discussions. If you orient your coding strand 5' left 3' right and orient your mRNA transcribing strand 5' left 3' right, you never need consider the template strand. The template strand is the strand the RNA polymerase is reading, deciding which base to add next.

Coding strand 5' left, 3' right. Transcribing 5' left, 3' right. Forget about the template strand. Draw it out until you get it; it pays dividends throughout this module.

General Structure of Genes

The other part that makes transcription hard is regulation. Because only a small piece of DNA is copied, the nucleus must have some way of knowing which part to copy, how often, and for how long. Each stretch of DNA that codes for a protein is called a **gene**.

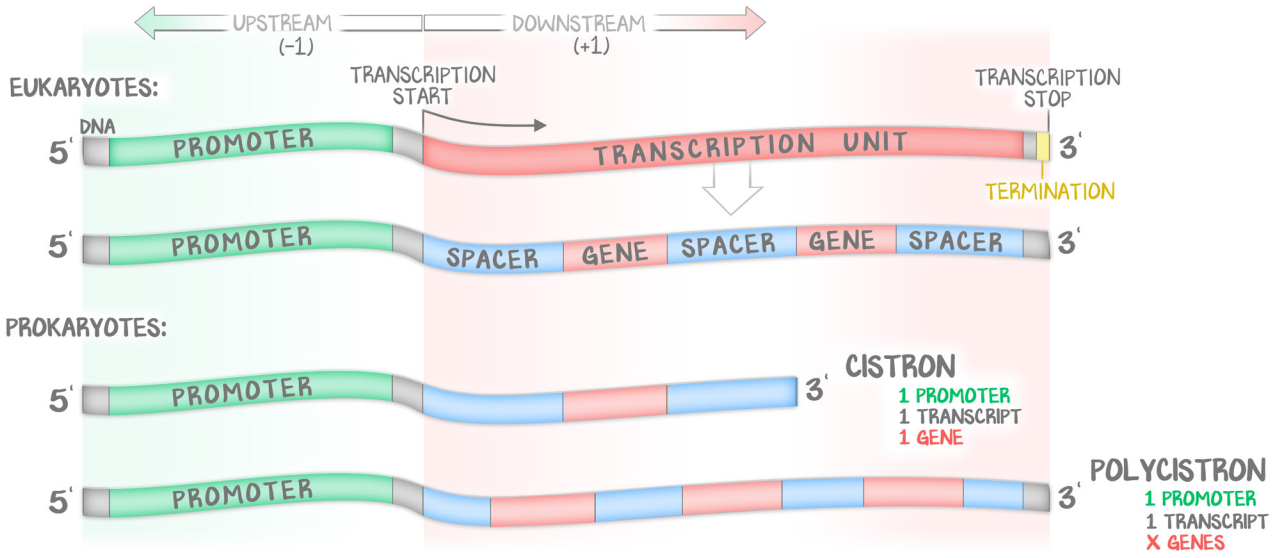

Figure 7.2: Spacer DNA and Cistrons
Spacer DNA does not code for a gene, and separates subsequent genes from one another. In a prokaryote, one promoter region with multiple genes is called a polycistron.

The general structure of a gene is to have a **promoter** that precedes the start of transcription. The **transcription unit** is "start" to "terminate." The **transcription unit** contains both **coding DNA** (the DNA that will actually become protein) and **spacer DNA** (DNA that doesn't become protein). Where the transcription unit begins is called the start; where the transcription unit ends is called the termination site.

Transcription has a start and a stop signal, and what gets transcribed is both coding and spacer DNA. Don't confuse this with the start and stop of translation, found at the beginning and the end of a coding region.

Anything **before start** is called **upstream**, and we mark it with a negative number, where the number is how far the base pair is from start. We refer to anything **after start** as **downstream**, and mark it with a positive number, where the number is how far the base pair is from start. The promoter is -10 to -100. The start is +1. Genes may be any number of nucleotides long.

Upstream (negative) and downstream (positive) are always in reference to the **coding strand** and therefore in reference to the **RNA-transcribing strand**. Anything that is negative is present in the coding-strand DNA but will be absent from the RNA-transcribing strand because RNA polymerase binds at the promoter region, already negative, and begins to transcribe at +1.

Eukaryotic DNA has to do more things with its DNA (multiple cell types in an organism) whereas prokaryotic DNA does not, so eukaryotic DNA has a lot of spacers, while prokaryotes have very little. The spacer DNA is not just wasted space. Instead, it's code that, if read with a different start position, could become gene DNA. Spacer DNA can be spliced as an intron (#9: *Eukaryotic Transcription*) or can be used to regulate other gene transcription as a binding site for transcription factors (#10: *Eukaryotic Transcription Regulation*). This has two consequences. First, a series of nucleotides can be read in multiple ways to produce many proteins out of the same code. Second, both prokaryotes (#8: *Prokaryotic Transcription*) and eukaryotes (#9: *Eukaryotic Transcription*) have to find ways of dealing with spacer DNA so that only the code actually needed for protein is read.

RNAs

mRNA is the **messenger RNA** and is the **RNA product of transcription**. It's the photocopy of the DNA meant to allow the code to leave the nucleus to the cytoplasm. mRNA contains coding and noncoding regions of nucleotides. mRNA is a string of nucleotides that'll present itself to ribosomes, which read the mRNA strand and **translate** it to amino acids via tRNA.

rRNA is the most abundant RNA in the cell. It's the RNA of **ribosomes**, the cellular machinery that reads the mRNA, matches the tRNA with the right amino acid to the mRNA, and adds the amino acid to the growing strand.

tRNA is the **transfer RNA**. It's used in **translation**. tRNA matches its anticodon with the mRNA codon (#12: *Intro to Translation*), informing the ribosome that the correct amino acid is ready to be attached to the growing chain. The amino-acid-and-tRNA combo enters the ribosome, where the ribosome takes the amino acid, joins it to the existing sequence, and ejects the now-free tRNA.

Figure 7.3: Types of RNA
The different ways "RNA" can be used in cells.

In eukaryotes, there's also **hnRNA** (pre-mRNA), which is the **primary transcription product**. Eukaryotic RNA polymerases make unprocessed hnRNA, which much undergo subsequent co- and post-transcriptional changes to be allowed out of the nucleus. Prokaryotic RNA polymerase makes the final mRNA (there is no processing). This is because there is no nucleus—prokaryotic mRNA can immediately begin translation, as both transcription and translation occur in the cytoplasm. Eukaryotic hnRNA must be tagged and dispatched from the nucleus into the cytoplasm for translation to begin.

In eukaryotes, there's also **snRNA** (small nuclear RNA) which, when combined with proteins, are called snRPS ("snurps"). snRPS are the RNA of spliceosomes, which allow introns to be removed from the primary transcription product, just one of three processing events (#9: *Eukaryotic Transcription*).

RNA Polymerases

RNA polymerase is **low-fidelity**. It has **no exonuclease activity** and therefore, by definition, has **no proofreading** feature. RNA polymerase lacks those abilities but is also liberated from any restrictions. By comparison, DNA polymerase is high-fidelity, but can't initiate replication without an RNA primer. RNA polymerases can **begin transcription without a primer** (in fact, the primer made during replication for DNA polymerase is the initial RNA sequence made by RNA primase, an RNA polymerase). And of course, this is **RNA**, so all structures will have a hydroxyl group at the 2-carbon position of the phosphate-pentose backbone.

RNA has A**U**GC, with Uracil being the only difference, replacing the Thymine in DNA's A**T**GC. To reiterate the point made earlier in this lesson, the coding strand and the mRNA transcribing strand will be identical except that the DNA-coding strand will have T's and the mRNA-transcribing strand will have U's. The coding and transcribing strands are NOT complementary to each other; they're both complementary to the template strand.

In prokaryotes, only **one RNA polymerase does everything**. In eukaryotes, each of three RNA polymerases does something. RNA polymerase I makes ribosomal RNA, the 40s and 60s subunits (50 and 30s is for prokaryotes). RNA polymerase II makes **all the stuff for transcription—**the **hnRNA** that becomes **mRNA** as well as the **snRNA** that processes that hnRNA. RNA polymerase III makes tRNA. The "gotcha" is that RNA polymerase III also makes the **5s ribosomal subunit** (which we won't hear of again outside this lesson).

RNA POLYMERASE I	RNA POLYMERASE II	RNA POLYMERASE III
rRNA—50s	**mRNA**	tRNA
rRNA—30s	**hnRNA**	**rRNA 5s**
	snRNA	

Table 7.1: RNA Polymerases and What They Make